IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
PATENT EXAMINING OPERATION

Applicant(s): Cesar Bandera et al.
Serial No.: 09/542,562 Confirmation No. 5137
Filed: 04/04/2000
For: DYNAMICALLY RECONFIGURABLE VISION SYSTEM
Examiner: Timothy J. Henn Art Unit: 2612
Atty Docket: ML-0495CV

DECLARATION UNDER 37 C.F.R. 1.131

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

Christopher Kramer declares that he is one of the inventors of the above-identified application, and that the invention of the application was conceived in the U.S. before the April 1, 1999 effective date of reference NASA Technology Brief JPL New Technology Report NPO-20866 (the reference), and that the Guang Yang, Chao Sun, Chris Wrigley, David Stack, and Bedabrata Pain paper entitled "Multi-acuity, Multi-window Dynamically Reconfigurable CMOS Imager for Real-Time Active Vision Systems" is a part of that NASA Technology Brief as evident from the common stamp 0464 2086 on the top of each page of the reference.

Enclosed is a copy of Draft No. 5 of the Provisional Application to which the above-identified application claims priority. Draft No. 5 dated March 26, 1999 was prepared by a consultant, Alexander E. Martens, using input from inventors of the Provisional Application. The consultant delivered Draft No. 5 electronically in an e-mail to inventor Cesar Bandera on March 30, 1999. Cesar Bandera on March 30, 1999 provided his approval of the Draft No. 5 by e-mail to Alexander E. Martens. Draft No. 5 thus shows conception of the invention prior to the April 1, 1999 effective date of the reference, and due diligence to the April 4, 1999 filing date of the Provisional Application No. 60/127,698 to which the above-identified application claims priority.

I further declare that all statements made here in of my own knowledge are true and that all statements made on information and belief are believed to be true; and, further, that the statement were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Respectfully submitted,

Date: March 5, 2004

Christopher Kramer

DRAFT #5

26 March, 1999

RECONFIGURABLE FOVEAL MACHINE VISION SYSTEM

Inventors: Cesar Bandera, 348 Parker Avenue, Buffalo, NY 14216
David J. Stack, 154 Central Avenue, Silver Creek, NY 14136
Christopher Kramer, 1514 Northwood Drive, Williamsville, NY 14221

Assignee: Amherst Systems, Inc. 30 Wilson Road, Buffalo, NY 14221

Filed:

ABSTRACT

A machine vision system is disclosed modeled after foveal vision of humans. The system includes a video camera that contains a two-dimensional photodetector pixel array and operates in conjunction with a host computer. For the purpose of optimizing the relevance of specific video information to the tasks being performed by the system, the system is designed to reconfigure electronically the resolution, size, shape, and focal plane position of the array's pixel geometry. The video information acquired in that manner is used for simultaneously detecting, identifying, and tracking multiple moving targets

FIELD OF INVENTION

Machine vision.

BACKGROUND OF INVENTION

Machine vision systems for detecting, identifying, and tracking targets must acquire and process large volumes of video data in real time. Imaging systems that employ uniform and constant spatial resolution throughout the entire field of view acquire much irrelevant information and thus burden valuable data processing and communication resources of the system. As a result, such systems are less effective and slower. To ensure adequate performance, such systems are made to be much more complex, expensive, large, and therefore are prone to failures. These constraints make constant resolution systems difficult or impossible to use where space, speed of response, and reliability are critical considerations, for instance, in defense applications. Two-dimensional photodetector arrays in systems of that type comprise pixels that are all of the same size and are hard-wired, with their geometry and organization remaining constant over time. The pixels are sequentially scanned, and the resulting video signal is then input into the host computer for processing.

Humans and other vertebrates have foveal vision that allows them to concurrently perform several tasks: survey a wide field of view at a low resolution for situation awareness and identification of features or targets of interest; track moving targets with

great accuracy; scan at high resolution these multiple targets of interest; communicate over channels with limited bandwidth (neurons) the information of interest to the computer (brain). Because high resolution imaging is limited to the fovea which is fixed in the center of the retina, the tracking of targets involves movement of the eyes and the head.

The machine vision system, the object of this invention, makes use of the biologically proven concept of foveal vision, and further improves on it by allowing simultaneous acquisition and tracking of multiple targets widely separated in the field of view, and performing these tasks without the need of mechanical tracking. Depending on the number of targets of interest, the system forms one or more high resolution windows within the low resolution field of view, each window containing a target of interest. Templates for identifying targets of consequence can be input into the system and stored to be used for automatic selection of such targets for review and tracking at high resolution. In addition, targets can be initially detected based on their movement, heat emission, shape, or color.

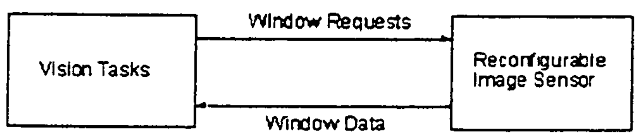

Figure A. Relationship Between Vision Functions and Dynamically-Reconfigurable Imager

The block diagram of Fig. A shows the essential components of the system and their interactions. Vision functions are performed by the host computer. The reconfigurable image sensor is a part of the video camera.

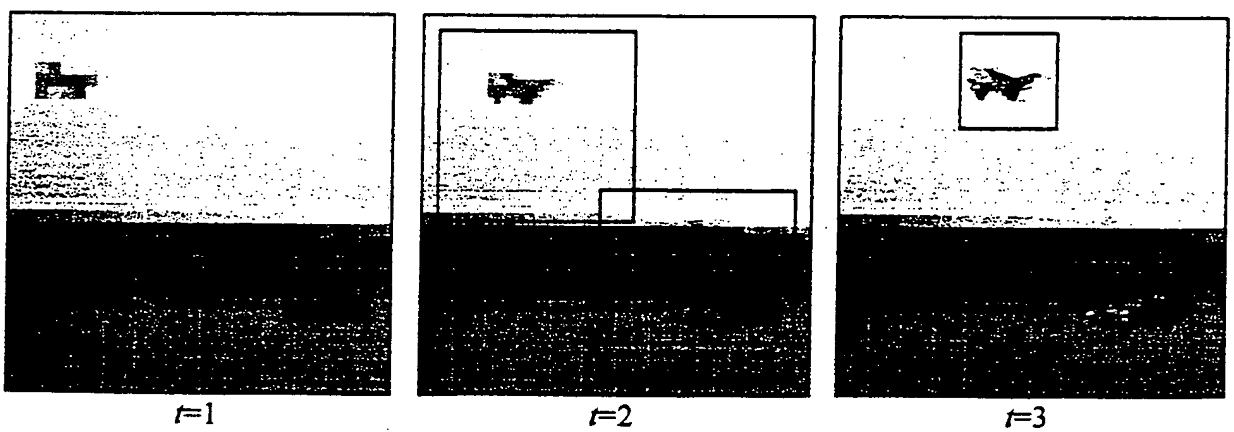

Figure B. An Example of Three-Window Foveal Imagery: Dual Target (a Plane and a Tank) Tracking plus Full Field of View Surveillance

Figure B illustrates the acquisition at low resolution of two targets of interest (t=1), the formation of a higher resolution window over each target (t=2), and the refinement of the two windows as they are reduced in size and increased in resolution (t=3). Note the effect

of increase in detail as resolution increases. Now, the two targets can be automatically tracked as they move. Should the movement take them out of the field of vision, the camera will be automatically pointed toward the target by a servo-controlled pointing mechanism. The size of the smallest resolving element can be automatically varied from the size of the individual pixel in the photodetector array to a superpixel, the size of the entire array.

Several previously described foveal vision systems employed techniques that were significantly different from those used in the system object of the present invention:

Photobit, Inc. of La Crescenta, CA, reported a vision system with variable resolution, in which groups of adjacent pixels are formed into superpixels, thus reducing the resolution. Rather than averaging the values of all pixels forming the superpixel, the value of such a superpixel is determined to equal that of an arbitrarily selected pixel within the superpixel; consequently, should the selected pixel have the value representing white in the image, though the rest of the image was dark gray, the entire superpixel would have a value corresponding to white, instead of dark gray. This creates aliasing (the generation of image artifacts from high spatial frequencies in the scene) that precludes target detection and recognition particularly in cluttered scenes.

The U.S. Pat. 5,626,871 discloses a multiresolution image sensor that outputs data representing a superpixel to a computer that controls the size of that superpixel. The computer extracts video data from the sensor one superpixel at a time, whereas our system extracts video data one frame containing multiple windows at a time, therefore saving time and reducing the amount of interaction between the camera and the computer. Furthermore, the values reporting the level of illumination of the superpixels generated by the system described in 5,626,871 are a function of pixel size in which this value is equal to the sum of the comprising pixels. The computer then has to normalize the superpixel values so that the image processing algorithms do not erroneously interpret a large superpixel to represent a bright scene region. This operation requires extra time and system memory, thus slowing down the image processing. Larger pixel values also require a wider dynamic range in the video communication circuits.

In addition, in that system the photodetector array does not operate in a "snapshot" mode, so that the exposure time can be different for different pixels after the system has been reconfigured, requiring the computer to carry out additional normalization of pixel values in order to avoid a mistake in interpreting the pixel values in the context of the image. Because the pixels in this patented system are exposed at different times, any motion in the field of view (due to camera and/or target motion) will introduce artifacts, such as target warping, that reduce the accuracy of the target classification. Furthermore, overexposed pixels in an image will appear brighter than properly exposed or underexposed pixels in the same image. This deficiency interferes with the commonly accepted method of interpretation of a pixel value in an image. We avoid this problem in our system by using an electronic shutter on the pixel array that ensures that or pixels are exposed simultaneously irrespective of their position, size, and illumination history. The pixel averaging technique of our invention does not require pixel value normalization by

the computer, nor a wider signal dynamic range; as a result, this feature reduces, respectively, the processing and communication requirements.

Fixed geometry multiresolution systems, widely reported in literature, have a two-dimensional photodetector array that has small size pixels in the center and larger size pixels in the periphery, closely simulating the anatomical structure of the retina. Such arrays are not reconfigurable, do not operate in a closed loop fashion with the computer, and require a pointing mechanism for gazing. Mechanical pointing suffers from instabilities and is much slower than adaptive reconfiguration of the pixel array topology as we use in our system. No pointing mechanism is required in our system as long as the target remains within the field of view of the camera. Because the topology of the array in a fixed geometry system cannot be changed on demand, some relevant regions of the scene may not be resolved adequately, as a result sacrificing system reliability and usefulness, while the irrelevant regions may be resolved too finely, hence reducing the efficiency of system resource utilization. Our reconfigurable foveal machine vision system described here does not suffer from these disadvantages because the topology of the pixel array can be modified frame-by-frame in response to the instantaneous conditions in the scene being imaged.

In the so-called pyramidal machine vision systems targets are detected using low resolution; these targets are then scanned using localized high-resolution windows in a fashion that could be also represented by Fig. A. The pixels that form the high resolution windows are not combined into superpixels directly in the pixel array, as we do in our system, but the information from individual pixels is combined in the computer so as to mimic superpixels. These techniques, however, require the processing of video from a uniform resolution camera to generate the pyramid data structure. As a result, the communications bandwidth and the video processing resources are not used any more efficiently than in a conventional uniform resolution camera.

The information on photodetector circuitry that performs snapshot imaging and pixel averaging has been published by Jet Propulsion Laboratory of NASA. This circuitry requires the host computer to send control signals to the sensor for the extraction of the signal from each superpixel. In our system the host computer needs to issue only one brief high-level configuration command per frame to the camera.

It is therefore an object of this invention to provide a machine vision system that is based on the concept of foveal vision.

A further object of this invention is to provide a machine vision system that includes a video camera and a host computer.

Another object of this invention is to provide a machine vision system in which said video camera and said host computer form a closed loop system.

Still further object of this invention is to provide a machine vision system that is capable of automatically changing the resolution at which images or parts of images are acquired.

Still another object of this invention is to provide a machine vision system in which said video camera includes a two-dimensional photodetector array comprised of a plurality of individual fixed size pixels.

An additional object of this invention is to provide a machine vision system in which the size of the image resolving element can be automatically varied under control of said host computer in selected locations of said two-dimensional photodetector array from the size of a single pixel to the size of the entire array, and to all resolving element sizes in-between.

Yet another object of this invention is to provide a machine vision system in which said two-dimensional photodetector array is an active pixel photodetector array.

A further object of this invention is to provide a machine vision system in which said array of larger size pixels is used to acquire images of the entire field of view of said video camera.

Still another object of this invention is to provide a machine vision system in which within the boundaries of said video frame one or more windows comprised of arrays of smaller size but a greater pixels density are generated any place wherever the target or targets of interest are located. The purpose of said windows is to resolve additional image details of said targets.

An additional object of this invention is to provide a machine vision system the frame rate of which can be proportionally faster as the number of pixels in each frame decrease.

Yet further object of this invention is to provide a machine vision system in which said photodetector array includes electronic circuitry simulating a function of an optical shutter such that when open said electronic shutter allows all said pixels to respond to the light entering said video camera, and when closed not to respond to said light.

A still further object of this invention is to provide a machine vision system in which the location, and size of said windows is automatically controlled in response to the location and size of said target.

Yet another object of this invention is to provide a machine vision system in which the resolution of said windows can progress in multiple steps from the lowest (that of a single pixel) to the highest (that of a single superpixel encompassing the entire field of view).

SUMMARY OF THE INVENTION

This invention is related to a machine vision system that is based on the concept of foveal vision. The system, object of this invention, comprises a video camera and a host computer, and is capable of automatically detecting, recognizing, and tracking targets of interest. In the process of doing so, said system forms the video frame windows of higher resolution corresponding to the location in the field of view of said targets. Multiple

windows to enclose said targets can be formed. Windows can overlap, and smaller windows supporting the tracking of targets can co-exist with a wide field-of-view window, supporting the detection of new targets.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a block diagram depicting the essential components of the system.

FIG. 2 is a more detailed block diagram of the system

FIG. 3 depicts the software architecture of the system

FIG. 4 illustrates the progress of target acquisition and tracking by forming within a video frame windows of higher resolution subtending the instantaneous locations of the target.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

FIG. 1 is a block diagram of the foveal vision system object of this invention. It shows that said system comprises two major modules: a camera **29** and a host computer **18**.

With reference to FIG. 2, the host computer **18** comprises a single processor **10** for execution of video processing and contains control software for the configuration of camera **29**. The host computer interface to the camera **29** comprises a digital interface **16** which sends camera configuration commands **19** from said host computer **18** to said camera **29**. Said digital interface **16** digitizes the resulting video signal stream **20** and stores it in the memory **11** of said host computer **18**. Said processor **10** is able to access said stored digitized video signal over the memory bus **17** and video display hardware **12**. The digital interface **16** to the camera **29** performs the following functions:

1. *Control signal generation*. The configuration commands, generated by the processor **10**, configure the camera **29**. Said configuration commands may be submitted on a frame-by-frame basis. A command can affect the configuration of one, more than one, or all of the windows in the subsequent frames. A signal conditioning circuit **25** converts the configuration commands into signals recognized by the internal circuitry **21** of the camera **29**

2. *Video data acquisition*. Since its imagery is frame-by-frame reconfigurable, the camera **29** output is not interfaced to a conventional analog input video frame grabber. Instead, the photodetector array **26** outputs an analog signal **27** that is digitized. The conversion from analog to digital video signal format may be performed either in signal processing circuitry **28** in a circuit that could be a part of the camera **29**, of the photodetector array **26**, or reside in the interface **16**. The digital video data are stored in the memory **11** for subsequent processing.

Housed within the camera **29** are a conditioning circuitry **25** for the configuration command signal, a reconfigurable photodetector array **26**, video signal processing circuitry **28**, a power supply **22** (power may also be supplied to the camera from an external source), and voltage regulation **24** and distribution **23** circuitry. The conditioned command signals **21**configure the photodetector array **26**, and the resulting video signals **27** are conditioned for transmission **20** to the host computer **18**.

Referring now to FIG. 3 that illustrates the architecture of the software, said software comprises Detection and Tracking Algorithms (DTA) module **30**, and a Real-Time Control and Data Acquisition (RT-CDAQ) module **33**. The DTA processes each image received from the camera **29** and subsequently responds with a window request **31** for new data. The Control Vector Generation module **34** converts these top-level window requests into camera configuration commands **36**, which are sent through the device driver of the digital interface **16** to the camera **29**. The Window Acquisition module **35** collects the received video information **37** via the digital interface's **16** device driver and buffers the requested imagery **32** before sending it to the DTA **30** for subsequent processing. The DTA **30** can then determine the next required data set and issue new window request(s).

FIG. 4 illustrates schematically the process of a moving target acquisition and tracking using first a low resolution window that occupies the entire frame (a), followed by the formation of a higher resolution window around the target (b), and, finally, by a small, very high resolution window containing the target (c). At this point the target can be automatically identified by comparing it to a stored template. This progression simulates human vision in that step (a) corresponds to peripheral vision, step (b) to perifoveal vision, and step (c) to foveal vision.

It is to be understood that the preceding descriptions are illustrative only and that changes can be made in said reconfigurable foveal machine vision system, subject of this invention, in its components, materials and elements, as well as in all other aspects of this invention discussed herein without departing from the scope of the invention as defined in the claims.

We claim:

1. A machine vision system utilizing the concept of foveal vision comprising:
a video camera including at least one photodetector array,
a host computer containing at least one digital data processor.

2. A machine vision system per claim 1 in which said photodetector array is an active pixel photo detector array.

3. A machine vision system per claim 1 in which said photodetector array contains a plurality of individual pixels that have fixed dimensions.

4. A machine vision system per claim 1 in which said video camera and said host computer form a closed loop interactive system.

5. A machine vision system per claim 3 which is capable of automatically under control of said host computer of changing the resolution of said video camera at which images or parts of images within a frame are acquired by automatically varying the number of said individual pixels that constitute a single image resolving element, said number of individual pixel can vary from a single pixel to all the pixels in said photodetector array..

6. A machine vision system per claim 5 in which the resolution of said photodetector array can be automatically changed in selected locations corresponding to the instantaneous location of the image of said target or targets without changing the resolution of the remainder of said photodetector array.

7. A machine vision system per claim 6 in which said resolution can be increased or decreased in multiple steps.

8. A machine vision system per claim 6 in which said local resolution changes are confined to windows the size and shape of which is automatically determined in said photodetector array under control of said host computer. A window as claimed here is a defined area within the boundaries of said photodetector array that contains the image of the target at any given time.

9. A machine vision system per claim 1 in which, in order to facilitate the initial acquisition of said target or targets from the entire field of view of said video camera, the boundaries of said window are congruent with the boundaries of the entire photodetector array.

10. A machine vision system per claim 8 in which said initial acquisition of said target or targets is accelerated by making the size of said resolving elements larger than the size of said individual pixels.

11. A machine vision system per claim 7 in which a plurality of said windows is generated corresponding to the number of said acquired targets.

12. A machine vision system per claim 1 in which the resolution in said windows is gradually increased as the size of said windows is decreased to facilitate the resolution of detail in said target or targets and the accuracy of said tracking of said target or targets.

13. A machine vision system per claim 2 in which said photodetector array includes electronic circuitry simulating the function of an optical shutter such that when open said electronic shutter allows all said pixels to respond to the light entering said video camera and when closed not to respond to said light.

14. A machine vision system ~~per claim~~ 12 ~~which is capable of~~ establishing the nature of said target by comparing ~~the~~ image of said target to templates stored in said host computer.

15. A machine vision system ~~per claim~~ 8 ~~in~~ which ~~the rate~~ at which said frames are generated by said video camera increases as the size of the windows and the resolution of said frames decreases.